International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 03 Mar. 2025, pp: 489-494 www.ijaem.net ISSN: 2395-5252

Explainable AI for Regulatory Compliance in Financial and Healthcare Sectors: A comprehensive review

Nikhil Gupta[1]
[1]*Independent Researcher/ Alumnus, IIT Bombay, India*
Corresponding Author: Nikhil Gupta

Date of Submission: 10-03-2025 Date of Acceptance: 20-03-2025

ABSTRACT: Artificial intelligence (AI) is rapidly transforming the financial and healthcare sectors, offering innovative solutions for tasks like fraud detection, credit scoring, diagnostics, and treatment planning. The use of AI in these sectors is increasing, driven by the need for automation, improved decision-making, and enhanced efficiency. However, the "black-box" nature of many AI models poses challenges for regulatory compliance. Regulations such as GDPR's "Right to Explanation" and the FDA's AI/ML guidelines demand transparency and accountability in AI systems. Explainable AI (XAI) addresses this challenge by providing insights into the decision-making processes of AI models, enabling stakeholders to understand, trust, and effectively govern these technologies. This review article explores the role of XAI in achieving regulatory compliance within the financial and healthcare sectors. We examine key XAI methods, including SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and counterfactual explanations, and discuss their applications in real-world scenarios. We analyze case studies, regulatory perspectives, and the challenges and future trends of XAI in ensuring responsible and compliant AI adoption.

KEYWORDS: Explainable AI, Regulatory Compliance, Financial Sector, Healthcare, Model Interpretability, XAI Techniques, AI Governance

I. INTRODUCTION

Artificial intelligence (AI) is revolutionizing various industries, with a particularly profound impact on the financial and healthcare sectors. The use of AI in these sectors is increasing, driven by the need for automation, improved decision-making, and enhanced efficiency. In finance, AI is being deployed for fraud detection, credit scoring, algorithmic trading, risk management, and regulatory compliance, such as anti-money laundering (AML) and Know Your Customer (KYC) regulations. In healthcare, AI is used for disease diagnosis, personalized medicine, drug discovery, and clinical decision support. While AI offers significant benefits, the opacity of many AI models, often referred to as "black boxes," raises concerns regarding their trustworthiness, accountability, and compliance with existing regulations. Regulatory bodies worldwide are increasingly focused on the responsible use of AI. The European Union's General Data Protection Regulation (GDPR) includes a "Right to Explanation," requiring organizations to provide meaningful information about the logic involved in automated decision-making processes that significantly affect individuals. In the United States, the Food and Drug Administration (FDA) is developing regulations for AI/ML-based medical devices, emphasizing the need for transparency and validation. Financial regulations, such as Basel III, require banks to have a thorough understanding of their risk models, including those based on AI. While capital adequacy is a key component of Basel III, the regulations also encompass broader aspects of risk management and supervision, promoting the stability and resilience of the financial system. These regulatory requirements highlight the need for Explainable AI (XAI), a field of AI that aims to make the decision-making processes of AI models more transparent and understandable to humans. XAI techniques provide insights into how AI models arrive at their conclusions, enabling stakeholders to assess their fairness, reliability, and compliance with relevant regulations..

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 03 Mar. 2025, pp: 489-494 www.ijaem.net ISSN: 2395-5252

II. THE NEED FOR EXPLAINABLE AI IN REGULATORY COMPLIANCE

The lack of explainability in AI models can lead to significant compliance challenges. For instance, a biased credit scoring algorithm that unfairly discriminates against certain demographic groups could violate fair lending laws. In healthcare, an opaque AI diagnostic tool that fails to provide clear justifications for its predictions could lead to misdiagnosis and jeopardize patient safety. XAI is crucial for building trust and confidence in AI systems, especially in sensitive domains like finance and healthcare, where decisions can have significant impacts on individuals' lives. Patients need to trust the diagnoses and treatment recommendations provided by AI-powered systems, while financial consumers need assurance that AI-driven decisions about loans or investments are fair and unbiased.

XAI plays a crucial role in addressing these challenges by:

- **Enabling Auditability:** XAI allows regulators and auditors to examine the decision logic of AI models, ensuring compliance with relevant laws and guidelines.
- **Promoting Fairness and Bias Mitigation:** By revealing potential biases in AI models, XAI facilitates the development of fairer and more equitable AI systems. For example, XAI can be used to detect and mitigate bias in loan applications by identifying features that disproportionately impact certain demographic groups. Similarly, in healthcare, XAI can help ensure that diagnostic algorithms do not exhibit bias based on factors like race or gender.
- **Building Trust and Confidence:** Explainable AI helps build trust among users and stakeholders by providing clear explanations for AI-driven decisions.

XAI benefits various stakeholders:

- **Regulators:** XAI provides the necessary transparency to enforce regulations and ensure responsible AI development.
- **Businesses:** XAI helps organizations understand and manage the risks associated with AI deployment, fostering trust with customers and partners.
- **End-users:** XAI empowers individuals to understand how AI systems affect them, promoting informed decision-making and accountability

III. METHODS AND APPROACHES IN EXPLAINABLE AI

XAI encompasses a variety of methods and approaches, each with its own strengths and weaknesses. The choice of XAI method depends on the specific application, the type of AI model being used, and the desired level of explainability. There is no one-size-fits-all solution, and practitioners need to carefully consider the trade-offs involved in selecting an appropriate XAI technique.

These methods can be broadly categorized as:

- **Post-hoc Techniques:** These techniques explain the behavior of an already trained AI model. Popular methods include:
  - **SHAP (SHapley Additive exPlanations):** SHAP values assign importance to each feature in a model's prediction by considering all possible feature combinations. This game-theoretic approach provides a unified measure of feature importance and helps understand how each feature contributes to the model's output. SHAP is particularly useful for complex models where individual feature contributions are difficult to isolate.
  - **LIME (Local Interpretable Model-agnostic Explanations):** LIME creates a local surrogate model around a specific prediction, approximating the complex black-box model with a simpler, interpretable model like a linear regression. This allows users to understand the model's behavior in the vicinity of a particular instance. LIME is advantageous for its model-agnostic nature, meaning it can be applied to any type of AI model.
  - **Counterfactual Explanations:** Counterfactual explanations identify the minimal changes to input features that would alter the model's prediction. For example, a counterfactual explanation for a loan rejection might indicate that increasing the applicant's credit score by a certain amount would lead to loan approval. Counterfactual explanations are valuable for providing actionable insights and understanding how to change the outcome of an AI-driven decision.

While post-hoc techniques offer valuable insights into the workings of black-box models, they also have limitations. They provide approximations of the model's behavior rather than a complete and accurate representation of the underlying decision-making process.

In contrast to post-hoc methods, intrinsically interpretable models are designed for transparency from the outset:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 03 Mar. 2025, pp: 489-494 www.ijaem.net ISSN: 2395-5252

- **Intrinsically Interpretable Models:** These models are inherently designed for interpretability. Examples include:
  - **Decision Trees:** Decision trees represent the decision-making process as a tree-like structure, with each node representing a decision based on a specific feature. The hierarchical structure of decision trees makes it easy to trace the path of decisions leading to a particular outcome.
  - **Rule-based Systems:** Rule-based systems use a set of explicit rules to make predictions, making the decision logic transparent and easily understandable. These rules can be readily inspected and understood by humans, providing clear insights into the model's reasoning.
  - **Linear Models:** Linear models, such as linear regression, have simple and interpretable relationships between input features and output predictions. The coefficients in a linear model directly indicate the weight assigned to each feature, making it straightforward to understand the relative importance of different factors.Intrinsically interpretable models offer high transparency but may not always achieve the same level of accuracy as more complex black-box models, especially for tasks involving high-dimensional data or intricate patterns.
- **Hybrid Approaches:** Hybrid approaches combine black-box models with explainable components. For example, a deep neural network can be coupled with a decision tree to provide both high accuracy and interpretability. This approach leverages the strengths of both types of models, allowing for accurate predictions while maintaining a degree of transparency.

To further enhance clarity and facilitate comparison, the different XAI methods are summarized in the following table:

Method	Description	Advantages	Disadvantages	Example Applications
SHAP (SHapley Additive exPlanations)	Assigns importance to each feature by considering all possible feature combinations.	Provides a unified measure of feature importance; considers global model behavior.	Can be computationally expensive for large datasets.	Credit scoring, fraud detection, medical diagnosis.
LIME (Local Interpretable Model-agnostic Explanations)	Creates a local surrogate model around a specific prediction.	Model-agnostic; provides local explanations.	May not accurately reflect the global behavior of the model.	Loan applications, image classification, customer churn prediction.
Counterfactual Explanations	Identifies minimal changes to input features that would alter the model's prediction.	Provides actionable insights; helps understand how to change the outcome.	Can be challenging to generate for complex models.	Credit risk assessment, medical treatment planning, fraud prevention.
Decision Trees	Represents the decision-making process as a tree-like structure.	Easy to understand and visualize; can handle both categorical and numerical data.	Prone to overfitting; may not capture complex relationships.	Customer segmentation, medical diagnosis, fraud detection.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 7, Issue 03 Mar. 2025, pp: 489-494 www.ijaem.net ISSN: 2395-5252

Rule-based Systems	Uses a set of explicit rules to make predictions.	Highly transparent; easy to audit and debug.	Can be difficult to maintain for complex tasks; may not generalize well.	Fraud detection, credit scoring, medical diagnosis.
Linear Models	Establishes a linear relationship between input features and output predictions.	Simple and interpretable; easy to understand feature importance.	Limited expressiveness; may not capture non-linear relationships.	Price prediction, risk assessment, demand forecasting.

IV. CASE STUDIES AND INDUSTRY APPLICATIONS

Financial Sector:

- **Credit Scoring:** XAI techniques are being used to explain credit scoring models, ensuring fairness and compliance with regulations like the Equal Credit Opportunity Act (ECOA). Companies like FICO and LendingClub are exploring XAI to provide transparent credit decisions to consumers. This helps consumers understand the factors that influenced their credit score and provides them with actionable insights to improve their creditworthiness.
- **Fraud Detection:** XAI can help financial institutions understand the factors contributing to fraud alerts, improving the accuracy and efficiency of anti-money laundering (AML) systems. By providing insights into the patterns and anomalies that trigger fraud alerts, XAI can help investigators prioritize cases and reduce false positives. XAI is also being applied to KYC compliance, helping financial institutions verify customer identities and assess risk profiles more effectively.

Healthcare Sector:

- **Disease Diagnosis:** XAI is being applied to medical imaging analysis, providing explanations for AI-driven diagnoses and assisting clinicians in making informed decisions. For example, XAI can highlight the specific regions of an image that contributed to the AI's diagnosis, helping radiologists validate the AI's findings and identify potential areas of concern.
- **Predictive Analytics for Treatment Plans:** XAI can help explain personalized treatment recommendations, enhancing patient trust and understanding of their care plans. By providing clear explanations for why a particular treatment is recommended, XAI can empower patients to actively participate in their healthcare decisions and improve adherence to treatment plans.

V. CHALLENGES AND FUTURE DIRECTIONS

Despite its promise, XAI faces several challenges:

- **Trade-offs between Accuracy and Interpretability:** Highly complex models often achieve higher accuracy but are more difficult to interpret. Finding the right balance between accuracy and explainability is crucial. In some cases, simpler, more interpretable models may be sufficient, while in others, the need for high accuracy may necessitate the use of more complex models with more sophisticated XAI techniques.
- **Regulatory Gaps and Evolving AI Governance Frameworks:** Existing regulations may not adequately address the nuances of XAI, and AI governance frameworks are still evolving. As AI technology advances, regulators need to keep pace with these developments and establish clear guidelines for the responsible use of XAI in different sectors.
- **Applying XAI in practice can be challenging due to the complexity of models and the need for domain expertise to interpret explanations.** Effectively implementing XAI requires collaboration between AI developers, domain experts (such as financial analysts or clinicians), and legal professionals to ensure that explanations are accurate, meaningful, and relevant to the specific application.

Future research directions in XAI include:

- Developing more robust and reliable XAI methods that can handle complex AI models.

- Establishing standardized metrics and evaluation frameworks for XAI. This includes developing quantitative measures of explainability and creating benchmarks to compare the performance of different XAI techniques.
- Addressing the ethical and societal implications of XAI.

VI. ETHICAL CONSIDERATIONS FOR XAI

While XAI offers significant benefits in terms of transparency and accountability, it also raises important ethical considerations. These include:

- **Risk of Manipulation:** Explanations can be manipulated to mislead users or conceal biases in AI systems. It is crucial to ensure that explanations are accurate, faithful to the model's behavior, and presented in a way that avoids manipulation or deception.
- **Privacy Violations:** XAI techniques may reveal sensitive information about individuals or data used to train the AI model. Privacy-preserving XAI methods are needed to protect sensitive data while providing meaningful explanations.
- **Responsible Disclosure of Explanations:** The level of detail and complexity of explanations should be tailored to the audience. Providing overly technical explanations to non-experts can be confusing and counterproductive, while overly simplistic explanations may not accurately reflect the model's behavior.

Addressing these ethical considerations is essential for ensuring that XAI is used responsibly and does not inadvertently exacerbate existing societal biases or create new harms.

VII. CONCLUSION

Explainable AI is essential for ensuring responsible and compliant AI adoption in the financial and healthcare sectors. XAI techniques provide the transparency needed to meet regulatory requirements, build trust with stakeholders, and mitigate potential risks associated with AI systems. By shedding light on the decision-making processes of AI models, XAI empowers regulators, businesses, and end-users to understand, trust, and effectively govern these transformative technologies. XAI is not just a technical challenge but also a crucial aspect of building trust and fostering collaboration between humans and AI systems.

The future of XAI in regulatory compliance hinges on several key factors:

- **Evolving Relationship with AI Governance Frameworks:** As AI governance frameworks mature, they need to incorporate XAI principles and provide clear guidance on how to achieve explainability in different contexts.
- **Interdisciplinary Collaboration:** Developing and deploying XAI responsibly requires collaboration between AI developers, legal experts, domain specialists, and ethicists to ensure that XAI systems are aligned with societal values and ethical principles.
- **Driving Innovation and Responsible AI Adoption:** XAI has the potential to drive innovation and responsible AI adoption across various sectors by promoting transparency, accountability, and trust in AI systems.

As AI continues to evolve, ongoing research and collaboration are crucial to address the challenges and unlock the full potential of XAI in promoting ethical and compliant AI development.

REFERENCES

[1]. Islam, M.U., Mozaharul Mottalib, M., Hassan, M., Alam, Z.I., Zobaed, S.M., Fazle Rabby, M. (2022). The Past, Present, and Prospective Future of XAI: A Comprehensive Review. In: Ahmed, M., Islam, S.R., Anwar, A., Moustafa, N., Pathan, AS.K. (eds) Explainable Artificial Intelligence for Cyber Security. Studies in Computational Intelligence, vol 1025. Springer, Cham. https://doi.org/10.1007/978-3-030-96630-0_1

[2]. A. Adadi and M. Berrada, "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)," in IEEE Access, vol. 6, pp. 52138-52160, 2018, doi: 10.1109/ACCESS.2018.2870052.

[3]. V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra and B. Sikdar, "A Review of Trustworthy and Explainable Artificial Intelligence (XAI)," in IEEE Access, vol. 11, pp. 78994-79015, 2023, doi: 10.1109/ACCESS.2023.3294569.

[4]. Yeo, W. J., Wihan, V. D. H., Mao, R., Cambria, E., Satapathy, R., & Mengaldo, G. (2023, September 21). A Comprehensive review on Financial Explainable AI. arXiv.org. https://arxiv.org/abs/2309.11960

[5]. D. Saraswat et al., "Explainable AI for Healthcare 5.0: Opportunities and Challenges," in IEEE Access, vol. 10, pp.

84486-84517, 2022, doi: 10.1109/ACCESS.2022.3197671.

[6]. K. Kalasampath, S. KN, S. Sajeev, S. S. Kuppa, K. Ajay and M. Angulakshmi, "A Literature Review on Applications of Explainable Artificial Intelligence (XAI)," in IEEE Access, doi: 10.1109/ACCESS.2025.3546681.

[7]. Zahra Sadeghi, Roohallah Alizadehsani, Mehmet Akif CIFCI, Samina Kausar, Rizwan Rehman, Priyakshi Mahanta, Pranjal Kumar Bora, Ammar Almasri, Rami S. Alkhawaldeh, Sadiq Hussain, Bilal Alatas, Afshin Shoeibi, Hossein Moosaei, Milan Hladík, Saeid Nahavandi, Panos M. Pardalos, A review of Explainable Artificial Intelligence in healthcare, Computers and Electrical Engineering, Volume 118, Part A, 2024, 109370, ISSN 0045-7906, https://doi.org/10.1016/j.compeleceng.2024.109370.

[8]. Ahuja, Ashutosh and Wairagade, Anant and Gupta, Nikhil, AIREA: An AI-Driven Optimization Framework for Intelligent Automation in Large-Scale Enterprise Systems (February 20, 2025). http://dx.doi.org/10.2139/ssrn.5160312

[9]. Gupta, N. (2025). Privacy-Preserving Generative AI Techniques in Smart Home Appliances: A Comprehensive Review.11. https://ijirt.org/publishedpaper/IJIRT173183_PAPER.pdf